AD

TR-1392

Part I

AUTOMATION OF THE ABC SYSTEM

Part I. Linguistic Problems and Outline of a Prototype Test

August 1968

U.S. ARMY MATERIEL COMMAND

HARRY DIAMOND LABORATORIES

WASHINGTON, D.C. 20438

AD

DA-2P020401A728
AMCMS Code: 5910.21.63071
HDL Proj: 01220

TR-1392
Part I

AUTOMATION OF THE ABC SYSTEM

Part I. Linguistic Problems and Outline of a Prototype Test

by
Berthold Altmann
Walter A. Riessler

August 1968

U.S. ARMY MATERIEL COMMAND
HARRY DIAMOND LABORATORIES
WASHINGTON, D.C. 20438

ABSTRACT

To advance the ABC system toward the automation of its retrieval and analytical input operations, linguistic problems were studied, and a prototype computerized retrieval test was conducted. A vector-type organization was imposed on the test collection.

An appropriate measuring tool was constructed and used (a) to evaluate a variety of system parameters (ca 50 test runs were required) and (b) to rate different systems that evolve from the basic ABC model.

The process of computerizing the standardization of ABC descriptors as well as the production of a comprehensive thesaurus (presenting terminology with associations and functions) are described and so are the methods prepared for progressive automation of the analytical effort in future test models.

ACKNOWLEDGEMENT

Dr. Werner Menden participated in the early phase of the study, especially in the preparation of the categories and vectors for the structure of the test collection. We owe a great debt of gratitude to our Technical Director, Mr. B. Horton, for giving us permission to employ HDL scientists and engineers in the task of converting the test collection for automatic retrieval operations and furthermore to the members of our professional community (whose names are listed in Appendix V) for the performance of the task.

Mr. Martin Shaver and Mr. David Marsh planned and programmed the actual computer operations. The mathematical results were evaluated by Professor Robert B. Helle , George Washington University and the entire study by Dr. Irving H. Sher, Director of R and D (Information Company of America, Philadelphia, Pa.), who offered also most useful suggestions. It is our particular duty to pay tribute to the late Dr. Samuel Alexander whose incessant line of questioning has motivated us to seek practical machinable solutions. His relentless critical questions forced us to seek practical solutions for the automation of producing standardized ABC descriptors.

We received from ARO under DA2P020401A728, ATLIS-Task Area 04 the funds which paid for the numerous computer programs necessary to perform the test.

CONTENTS

FORM 1473, Last Page, Part III

A. INTRODUCTION

Two years ago we completed a manual retrieval test to identify and assess the characteristics of the first-generation ABC system. A mathematical (statistical) model[1] was used to calculate the standard deviation for relevance and recall to a high level of confidence. In general, it corroborated the conclusions of a preliminary report on the test.[2] These results indicated that a relevance ratio of 86-87 percent was characteristic of the system under the experimental conditions specified. However, an equally valid recall figure could not be established on a sound statistical basis because of the difficulty in specifying the total number of documents in the collection relevant to the query and because the volunteer retrieval operators typically stopped searching after withdrawing an average of only two documents per retrieval run. With a mean of a pooled total of eight unique titles responsive to each query, the maximum recall possible per individual was 25 percent; actual computations yielded approximately 22 percent.

Another matter discussed was the nature of the interdependence of relevance and recall and their validity as measurements of retrieval performance.

Our experience with the test and with day to day operations confirms the often reported observation that the idea of relevance is peculiar with each individual investigator.

Work on the second-generation ABC system has continued[3] to improve the system and to obtain the following broad objectives[4]:

1) streamlining the syntactical structure of the ABC descriptor

2) achieving a high degree of consistency in the descriptors' terminology of descriptors and sequence of concepts

3) improving the production of the ABC descriptor dictionary to meet the actual requirements of subject specialists (in particular for manual retrieval)

4) automation of the retrieval operation and

5) gradual automation of all input (or processing) activities

The research within our organization was conducted to gain a better understanding of semantic and syntactical properties of the ABC descriptor, to design a method for their automated construction, to automate the retrieval operations, to design objective testing and

evaluation methods appropriate to study the important elements of information retrieval, and to compare different retrieval runs and systems for use in developing and testing the second-generation system. A contract was let to produce a much larger collection, and the most important requisite computer programs were completed according to a detailed and comprehensive systems analysis.[5]

This report will therefore deal with the following aspects:

1) the linguistic problem of information retrieval

2) the preparation of an automated prototype retrieval test

3) the model and the analysis of the automatic retrieval test, and in particular, the development of an independent measuring tool

4) the current preparations for testing the growing new collection and the on-line (real-time) retrieval capability of the system and

5) the feasibility of automating the operations and services of an information office.

B. LINGUISTICS

a. The Linguistic Problem of Information Retrieval

For an identification of the role and place that language research must occupy in the development of modern retrieval systems we can call attention to the conclusions of the "DOD user needs study," an inquiry based on a representative and detailed sample of reactions from the scientific and technical community. The investigators in discussing one of their observations, namely the disregard of the established formal information system, offered among others the explanation that the formal system apparently does not provide the features desired by the user, particularly the features of convenience, responsiveness, and the ability to conduct a dialogue with the system.[6] The convenience[7] and the responsiveness scholars and scientists expect from a modern retrieval system have been more precisely defined by Leimkuhler and Neville in their statement: "The task of the specialized libraries and information services is not simply to amass all the research material in their particular field, but to organize and index it in such a way that questions about minute and precise topics can be quickly answered."[8] By implication the authors of both publications acknowledge that the basic problem of adequate reference or information services does not primarily depend on the development of new hardware, better computers and improved communication equipment, but on methods and procedures that will permit the "dialogue" between the

investigator himself and the stored analytical information. Scholars and scientists want to find their own answers to particular problems and they want to find them in the form, to the extent and within the time frame, suitable to their current task.

One key difficulty facing the designer of a modern and acceptable information system is a communication and therefore a language problem. First of all, complex subject matter must be organized for multiple approaches, while materials easily identified by one, two or three terms (or subject headings) can and should be serviced by the conventional catalog as the most economical and effective retrieval tool.

Second, whatever their vision concerning a future automated retrieval system may be, most documentalists agree that the ideal system should permit direct communications of the investigator with the organized stored information.[9] The ideal procedure to follow will be the same one used when locating information in a published bibliography, or in a collection of books and journals, that is a form of more or less systematic browsing and a search increasing in momentum and precision while one pursues his effort. For providing an accurate statement of his information requirements, the reader must be far advanced in his study. In the initial phases of his investigation, however, his approach, if not his objectives, will sometimes be hazy because he frequently starts with hunches or doubts. During such periods of uncertainty, pieces of relevant information are most urgently needed and most sincerely appreciated. Therefore systems organized to serve a scholarly community must provide the capability for the investigator himself to perform screening or searching in order that he may profit from the interplay of modern on-line (real-time) communications. For the required interface, complex information must be presented in clear and commonly understandable formulation.

Third, the designer of a system must solve a translation problem. The information facility stores and processes books, journal articles and reports reflecting the diversity of vocabulary, syntax, and style of their individual authors. Language acquired by social processes, training, education, reading and discussions differs not only from group to group and from discipline to discipline because of distinct historical, cultural or professional environments, but also from individual to individual because of the countless ways similar ideas and concepts can be expressed.

The language of the users of the system, probably the most important but sometimes most neglected group, enters as formulation of requirements reflecting the users' ability, personal knowledge and understanding of his problem and of the system. Nevertheless in each

instance the user expects to acquire suitable answers and to find those small subsets of the collection that contain the most appropriate available answers.

It is the task of the organizers, the indexers and analysts, to facilitate retrieval of papers and documents on specific and closely related subjects despite the great variety in style and terminology authors as well as users employ in their communications.

b. Coordinate-Index Type Systems

The importance of the linguistic problem has been generally recognized. To enhance communications in the two-way channel connecting users and stored information, the generally applied, coordinate-index systems are believed to require the pre-establishment of a standard term dictionary with definitions and cross references to inform the indexer as well as the retrieval operator about the accepted pertinent terminology.

Despite large and still rapidly growing investments in these term and dictionary-based systems, one should not be afraid to pause and take stock of the inherent deficiencies that are commonly known, but in general discounted.

If we assume that an appropriate, comprehensive balanced dictionary of consistent standard terms can be produced, updating the dictionary will involve discussions, controversies and undesirable time delays.

More critical still is another requirement stressed recently by Cyril Cleverdon, the requirement of cross-referencing the terminology in order to establish the hierarchical, whole-part, broad-narrow or other types of relatedness. From a term such as "insulation" the thesaurus should refer to all materials used as insulators as well as to the different types of insulation. From "supersonic flow" it should lead to the related subjects of shock waves and shock tubes, and from "vibration" to "elasticity".[10] Although these cross-references are indubitably the prerequisites of satisfactory retrieval operations, it is questionable whether any subject specialist or any group of subject specialists will be able to construct a complete network of all known interdependencies or to predict all other possible interconnections, the innumerable relations, not only of super- or subordination, but of brother, sister, niece and nephew associations that may parallel or link the different concepts presented in literature or speeches. Under these conditions it is not difficult to find gaps and inconsistencies in modern thesauri. But how could it be otherwise? The difficulty of providing the cross references is a basic deficiency of the coordinate-index type systems.

Another premise on which these systems are based, the inter-indexer consistency, turns out to be falling short of the desired reliability. It is discouraging that some analyses have shown inconsistencies of over 80 percent when experienced indexers picked indexing terminology for identical documents from one specific thesaurus.[11] Such discrepancies cannot be avoided whenever indexers must assemble terms descriptive of the author's thesis and translate these assembled terms individually into "equivalent" expressions of the thesaurus.

Various investigators point also to the critical number of indexing terms required for optimum retrieval operations.[12] They could refer to a retrieval run where the query expressed by one combination of indexing terms will also cause the withdrawal of all documents indexed by losely related terms.

Although greater indexing consistency can be expected from automatic processing, one must take into consideration that symbols or words rarely represent one specific meaning, but as a rule project a variety of meanings and stand for up to 60 and more different connotations according to the context of terms in phrases, sentences and paragraphs.

As a writer or speaker can fill one symbol with a great variety of meanings, so can one and the same concept be expressed in numerous different ways,[13] e.g.; as noun-noun, adjective-noun, prepositional noun phrases, as infinitive or participle constructions, or as a subordinate or a main clause. Under these circumstances consistency cannot be accomplished by a simple look-up as much as we may wish that it could be done.

In a period of rapid advances in science and technology, the corresponding semantic changes should be more seriously considered than they presently are.

Quite unsatisfactory appears to be the lack of syntax[14], [15],[16] when isolated terms and phrases must be combined to represent the content of involved papers or queries.

These deficiencies are further compounded when a documentalist must interpret the queries in preparation for the computer programs without having the investigator enter the process before he receives the papers withdrawn from the collection for his use and evaluation.

As a rule an operating system of this type permits the investigator to determine the non-desired materials he has obtained, but prevents him from adequately estimating the number of pertinent documents the system has failed to recall in a particular run and at a given cut-off.

All these and other deficiencies lumped together presumably result in the linear inverse relationship of the commonly used performance measures, the so-called relevance (or pertinence) and recall ratios. The fatalistic attitude with which unsatisfactory systems were accepted has probably hampered the development of adequate testing methods and procedures which can only be refined in connection with progressively perfected systems. On the other hand it has impelled the more activistic documentalists to urge for greater improvements of coordinate index systems or to replace them with entirely different ones.

c. The ABC Method

Although our investigations imply that the introduction of sophisticated but practical processes could make coordinate-type indexing acceptable to the community of scholars and scientists, the proof can be offered only by future practical, unbiased tests. The understanding of the characteristics of a useful and economical information retrieval system was gained through our conceptual approach to the problem. In the approach by concept (ABC method) system, syntactical units were introduced. The content of the sentence was objectivized and brought to a stand-still by transforming the verb into the governing noun; and a KWIC-type computer program improved to display 450-character long structures in a more readable format (See Appendix II p. 73) provided for an economic and comprehensive system of cross-references (as far as the accepted standardized terminology is concerned).

The manual test (the results briefly mentioned above) yielded a relatively high and consistent relevance ratio, but was not conclusive with respect to recall.

When we resumed our development program after the test, we faced the following difficult linguistic problems:

1. to prevent the KWIC-type ABC-Dictionary of complex descriptions from growing into unmanageable proportions;

2. to apply syntagmas (apparently one-time formulations, see p. 53f below)and, nevertheless, organize information economically;

3. to standardize syntax, a task we failed to solve through standard operating procedures and human editors;

4. to identify all possible dependencies and correlations beyond the relatively intricate system of cross references the KWIC computer program produces;

5. to automate not only the generation of ABC descriptors and of the editorial effort, but also, gradually the entire analytical effort.

d. General Considerations

Any information retrieval system that operates with natural language either in form of combinations of index terms or in form of syntactical self-explanatory structures must be based on underlying principles that govern the relationship between meaning and language, or thinking and speaking, and that link linguistic symbols denoting related subject matter into a great variety of stronger or weaker associations. Only if we succeed in this effort can we expect to break the barrier formed by the unending wealth of our language, and provide the scholar and the scientist with the opportunity to browse and the complete freedom to choose.

Because the problem of storage and retrieval systems is a problem of organizing messages it is proper to identify linguistic semantic and syntactical theories that can be readily adapted by both the analyst (indexer) and the user. A sound structure can result when we build a system on assumptions thoroughly tested for adaptability and for reliability. In our current and future studies and tests, we therefore must seek answers to a number of fundamental questions: What combinations of words, terms or phrases, and strings of phrases will communicate meaningful concepts that are unlikely to be misunderstood?

Can we construct logical or hierarchical schemes for the efficient, continuous, comprehensive, unambiguous organization of literature? Is there a positive relationship between the frequency of occurrence of individual words or co-occurrence of words in descriptors and the content they represent? Are services performed at a low cost truly cheap if they do not adequately meet the fundamental requirements of the customer for whom they have been designed and if they are not capable of leading toward automation of the retrieval and the analytical input operations? In other words what is the cost effectiveness of the models? Can we build a practical system if we omit provisions for continuous checks, controls, and adjustments of the "index" language? Can we design a system that can organically evolve in response to advances in knowledge and science, a system easily adaptable to the changes we must expect? Finally can we arrive at practical approaches and solutions to automate the input operations?

As in other studies published in this series we do not deal with computer logic and computer language.

Only the problems of designing, improving and automating a modern retrieval system are treated and this as far as they are of a semantic and "thinking psychological"* nature.

We must also emphasize that the following discussions will be limited to a few general observations and in particular to some remarks that may supplement the excellent survey of Eric de Grolier[17] who has assembled and described a very comprehensive body of information concerning the basic problems such as the difference between machine translations and information retrieval, and the particular relations to philosophy (logic) and philology; and furthermore, has subjected the commonly known theories and systems to critical and most convincing evaluations.[18]

e. Definitions

1. Terminology used to Describe the Natural Language Approaches of the ABC Storage and Retrieval System

Words or terms are language symbols that as a rule do express a number of meanings depending on the context in which they are placed.

Phrases are modified terms, or combinations of words used to identify one specific object or entity and therefore to form particular concepts. The combination may be noun-noun, adjective-noun, or preposition-noun.

One string of phrases, an ABC descriptor, circumscribes one problem or one task having one specific objective.

Usually, 3 to 5 ABC descriptors logically arranged will represent the content of one study, one research effort, one report or one publication; they form a summary of the different specific aspects treated in one bibliographic unit.

Categories are general subject groupings developed for the organization of sets of ABC descriptors or component phrases thereof.

The categories currently used have been the result of long-time observations and therefore reflect (a) HDL fields of interests (discipline-orientation), (b) HDL tasks or activities (mission-orientation) including characteristics, properties, or parameters, and (c) particular types or purposes and uses of pertinent publications with respect to form, level of difficulty, work phases described, etc.

*This term coined by Hoenigswald will be defined later.

Any detailed analysis implies the imposition of a rationale of classification. However, the never ending task of the documentalist is to redefine the rationale of categories and groups of categories to meet not only the changing requirements of his clientele, but also to organize the advances and progressive specialization of the disciplines represented in his collection. The categories used in the HDL system are more specific than in other systems and intentionally overlapping for several reasons: (1) to permit the coverage of marginal fields in more than one of the ABC category dictionaries and word lists produced for subject specialists, (2) to automate the organization of vectors for one particular retrieval method; and (3) to explore additional computer applications to clumping and similar processes the practicability of which must be explored.

The following explanations may also assist in clarifying our present position which will be subjected to stringent tests. (1) It is not possible to design categories which are completely exclusive of each other. (2) As we experienced in our prototype test which we are about to describe, an increase in the numbers of substantial categories assigned to the average document in the test collection improves the retrieval capacity of the system. If we use broader, overlapping categories, we have a greater chance of applying the average category more frequently and of obtaining a larger number of measureable components for the documents organized by a vector-type retrieval method. A larger overlap of categories appears to assure greater precision of identifying or locating the analyzed documents in a multidimensional space. However, an adequate balance must be established between two extremes, the categories so specific that they can be rarely applied and the categories so broad that they can be assigned to a majority of documents and therefore become useless for classification and retrieval.

Clusters in the ABC dictionary denote groups of ABC descriptors having one content-bearing term (keyword) in common. Arranged alphabetically, clusters define the keyword by providing the context of the different assembled descriptors. Guides to a keyword are the category term dictionaries or the content-bearing terms encountered in any of the other clusters or descriptors.

The structured abstract is a computer product, a mechanical transformation and printout of the subject analysis performed by the subject specialist. The specialist using a standard worksheet (ChartD.) identifies the pertinent, valuable information of the document, and organizes the content of a paper under such broad classes as: hardware, disciplines, production methods, functions and operations, influences, environmental factors, etc. The worksheet

also assists the linguistic editor in developing and applying consistent, unambiguous descriptor terminology; and moreover will in combination with a computer program improve the question-answering capability of the system.

2. Document versus Information Retrieval

As a research establishment responsible for the support of current programs and projects, the information office has to operate not only within certain time limitations but also according to particular quality standards. Seldom does there arise a requirement for a bibliographic compilation covering entire subject areas, because as a rule no one has time for long searches and intensive studies. The engineers as well as the scientists seek immediate answers to problems with which they are confronted. The objective is most often well defined, the operational environment of a required device or system is also generally known but the approaches that could be tried or the methods that should be applied must be determined. Difficulties may arise when seeking the pertinent, and if possible the most useful, information because the terminology used by the investigator in stating his problem may not be appropriate when the materials or components he is seeking or the method best suited to meet his requirement may have been previously studied or used in connection with an entirely different task.

Moreover it is the responsibility of an information office in a research installation to locate and withdraw a very limited number of documents or papers containing the specific information and to assist the investigator in answering specific and often very complex questions.[19]

The ABC system differs from other storage and retrieval systems in that it has not been designed to operate from a structured input. However, in one of its later versions the system will exhibit a capability of answering questions about particular methods used under precisely stated conditions, about devices, sub-systems or systems possessing particular properties and characteristics and about principles and designs studied or tested for particular application--in short, information in which a chemist, a physicist, an electronics engineer or an operations analyst might have a common interest despite their different approaches. In brief, the system is designed to identify the very few documents that contain the most appropriate answers.

f. Semantic Theory

The German semanticist Jost Trier[30] added a new dimension to our understanding of the relationship of language and meaning when he

concentrated his studies upon large cross sections of particular conceptual entities at different sequential periods and thereby introduced the concepts "word field" and "dynamic force" into semantic research. With every word used in our communications, Trier points out, we bring to our mind and to the listener's or reader's attention a number of meanings and words having relationships of varying distance from the terms we have chosen to use. All these related words and terms have a common ground; they form one well-structured organization, composed of layers of linguistic symbols called "word fields."

The words in one field are mutually dependent and receive their particular meaning or their conceptual content from the complete field in which they exist.

According to Trier, no a priori clearly defined concept is assigned to a particular word. On the contrary, the words assembled in one field receive their individual share of meaning by continuous mutual re-delineations and re-adjustments that take place within their field. The thinking human being throws a net of words over what is mere intuition, subconscious perception or guess, in order to "catch" a concept by comprehension and translate it into defined terminology. Language does not necessarily mirror reality, but creates intellectual symbols to facilitate the understanding of realities.

Trier illustrates his theory with his historic approach; that is by comparing closely related word fields from different periods. Because each word is meaningful only as a member of a given field, and in its relation to or distinction from all other members, any addition, substraction, or any shifts of word characteristics must cause disturbances and vacillations that will continue until the symbols again are adjusted to each other in the re-established conceptual complex. As the words are quasi-stable and are defined merely within the temporary configuration of a given field, so are the word fields themselves subject to readjustments at any time when one of the neighboring fields is given even a slightly different meaning.

In this steadily changing structure the individual plays an important role because only in him and through him does language find its realization. However, his role is limited in so far as it depends upon his reaction to a set of very strong, culturally and socially bound traditions.

With Trier's concepts of a living and dynamic language we can recognize one important element of our communication problem. In a period of extensive research and of steadily advancing and increasing knowledge new thoughts and new concepts will arise that demand either new symbols or the re-allocation of meaning within the existing "word

fields." Through the incessant influx and the continuing adjustments necessitated by the growth of ideas and knowledge, certain fields must be greatly disturbed and will seldom maintain a long-time balance.

Under the influence of progressive individualism, language as a tool of concept realization must yield to numerous demands of groups and individuals in modern society, and accept changes at a steadily accelerated speed.

Trier places his emphasis on semantics or words and their changing meanings within given fields, and on the comparison of historical cross-sections. Only in passing does he mention the "syntactical fields" which like "word fields" form a system, apparently a separate one and find their realization also in thinking and communicating individuals.[21]

Trier's limitation can be traced to practical considerations. The comprehension of the semantic role of syntax was provided more readily by the entirely analytical approach of a philosopher.

g. The American Psychological Approach

Quillian's[22] memory model and word associations grouped in different planes could be considered a transformation of Jost Trier's semantic fields("word fields")to permit automation. Each of Quillian's planes contains one "type node" surrounded by all the terms (or "token nodes") that contribute to its meaning. A multidimensional network is established in that most token nodes are also type nodes in planes of their own where they are defined by their respective token nodes. Indeed, Quillian's model of a semantic memory in which "every word is the patriarch of its own separate hierarchy," with the different planes related variously to each other without an absolute hierarchy, is descriptive of the ABC dictionary particularly in the first generation format. The alphabetization introduced to construct the clusters of ABC descriptors may generate groupings of inferior quality when compared with those formed by the free associations around the type nodes. This deterioration is to some extent compensated by the variety of keywords which are encountered in the different descriptors of the cluster and should be used as guides to other related subject matter, which in turn should lead to additional information so that the entire subject matter is made available in a number of sequential steps. Since the surrogate of each document that was selected for its quality and pertinence is presented in standard terminology (and in time also in standard sequence of notions and in standard syntax), the co-occurrence of terms in descriptors and clusters cannot be regarded as a matter of chance but must indicate a relatedness of

information. Thus when an investigator follows relevant terms and examines their respective clusters he is able, in 5 to 10 minutes, to retrieve the related literature. Often this will include subjects he may have forgotten or failed to consider when he started his investigation.

The retrieval with the ABC dictionaries turns out to be a search process more similar to one based on word fields or planes than may be apparent at the first glance. The completely cross-referenced listings of the ABC dictionaries are prepared and updated by an integrated computer operation.*

h. Syntax and Semantics, The Thinking-Psychological Approach

For the most provocative ideas concerning the mutual relations of logic and language, thought and expression, word and sentence, we are deeply indebted to the work of Richard Hoenigswald.[23] His analyses and definitions offer an understanding of the mental processes leading scholars to the formulation and confirmation of truth. His work provides an insight into the thinking operations and into what makes an activity productive or creative, and by implication, into the principles of storage and retrieval systems that meet the basic and the functional requirements for the support of creative research.

The thinking process is an intentional effort of an individual to establish a personal relationship to "something." He may start with knowing something that he is not fully conscious of and something that remains uncertain and vague until it is adequately defined, so that it can be expressed in terms of language.

Language thus has in an inextricable association with perception and cognition. Sense or meaning is basically verbal, and can be manifest only in worded form.

Language raises a personal act of perception to the level of knowledge, possible truth, and truth, and in this respect, has a functional character. It is a producer of thoughts as well as an equivalent of thinking;[24] and in so far as one learns the structure and use of a language, it is a product of society and civilization.

A thinking process in its most primitive form is presentable as "I have something" or, as we mentioned above, the smallest thinking event represents a relationship between an individual and a "something." It is, therefore, evident that the linguistic equivalent of this relationship is the sentence, and although the shortest

*See Appendix I.

linguistic unit capable of expressing or transmitting a thought may sometimes lack the grammatical form of a sentence, it will always be representative of and understandable as a sentence.

Especially in modern languages, words carry a great variety of meanings. A specific function, a connotation, a value is assigned to them by their position in a given sentence and sometimes by their context in a paragraph or a longer exposition. Syntax and semantics are inseparable for the philosopher because the meaning of words and their syntactical formations are as closely interrelated as logic and language.

In this highly condensed review of a philosopher's deep and complex analysis we will profit from following him a step further. The "something" of perception cannot exist in isolation. It is perceived as being related and if so observed as a complexity. The individual who perceives the "something" and the "complex" related thereto could not perceive the something without perceiving the "complex."

Whenever a thought relates to two or more "somethings" which in turn can be related to other somethings, more and more complex thinking events occur. These more complex thoughts representable by sentences must nevertheless be capable of being expressed within another sentence. With sentences belonging to a higher logical level because they characterize broader and still broader sectors of research and knowledge, we arrive at the essence of knowledge where everything is one and one is everything. Hoenigswald's analysis of the thinking process, based merely on theoretical, logical considerations, is also a viewpoint of reality. Regarding this he introduced his concept "präsenz," the capability of the human being to have present in his conscious mind a multiplicity of thoughts or thinking events at one brief "present" moment. This compression of previous thinking events into one thought incident enables man to compare, analyze and judge. It makes the human being a productive, sometimes a creative factor in life.

If we search for a physiological explanation of this phenomenon that obviously does not operate with the physical time concept of past-present-future, we gain more insight from the model of the brain and its function that biochemists such as Holger Hyden have proposed. According to this theory, information is stored and retrieved through biochemical processes that modify proteins and ribonucleic acid molecules within the synapses--the 4-10,000 junctions that tie the 10^{10} neutrons of the human brain into one complex neural network and react to message impulses by emission of requested stored substance information.[25] These experiments and investigations have not been completed.

An obvious conclusion of these discussions is that one cannot generate meaningful and active communications merely on the basis of a standardized vocabulary, however valuable and necessary that may be. Information can hope to be unambiguously identified for efficient retrieval operations only in context or within syntactical structures or sentences.[26] While the standardization of terminology and phrases may well lie within current machine and programming capabilities, the production of a standardized syntax could, according to various documentalists such as Kasher,[27] turn out to be one of the enterprises that are impossible of execution.

i. The Thinking Machine

In the Western world the ideal of a thinking machine and of conversations between a scholar and a computer have remained subjects of serious discussion.* On the other hand a number of Marxist scientists have accepted the doctrines of the Neo-Kantian Philosopher Hoenigswald.[28]

---

*For a physicist as Dean E. Wooldridge [The machinery of the brain. 1963] the brain resembles an advanced computer system. Although he agrees that most complex processes are required to sustain intellectual activities he describes the generating machinery of human thoughts as a complex neuronal network with storage areas, input and output couplers, filters, control, switching and feedback mechanisms.

When the continuously received impressions are directed to still vacant storage areas, an "automatic-pattern-interconnection principle" produces interconnections between the various sensory patterns and activates one memory pattern. Where "patterns containing similar sensory content" are stored, the threshold of the particular areas is lowered, while "an inhibiting mechanism...filters only one chain of recollections at a time out of the memory store."

Explanations of this type do not solve a number of baffling problems. The machine of today cannot duplicate the peculiar thinking phenomena because computer operations take place in physical time, they are sequential, they process, arrange, or match one object after another; they cannot produce as one thinking event, the awareness of an entire segment of knowledge, or of many segments belonging to different disciplines. Computers of today and of the near future can at best process, select and organize information for utilization, evaluation and creative synthesis by human beings.

The question must also be raised how and to which extent the suggested electronic and mechanical operations can convincingly

They allege that the position of Marxist philosophy and psychology is consistent with the argument that thinking is the cognitive reflection of the surrounding world in the human brain, and that it is manifest in the form of concepts, judgments and inferences. "Thinking is inseparably linked to consciousness and language....The machine computer has no concepts, including the concept "I"., and without the concept "I" the distinction between subject and object of reflection is not possible...machines do not think...,". Machines may be useful to "model various psychic processes, including thinking," but not "to uncover the qualitative, specific character of these different phenomena."

However, the USSR scientists concede that the creation of "brain matter" or of "an artificial creature" should not be considered an insoluble problem. Such developments would be completely different from "devising a thinking machine."

j. Discussion

The information office in a research establishment is responsible for supporting current programs and projects with information according to particular quality standards. These standards have apparently not yet been met according to the interviews by the Auerbach Corporation of the DOD professional community[39] and according to the linguistic analysis we have presented on the preceding pages.

Excluding the quality and pertinence of acquired information and the determination of such characteristics by qualified analysts, the mismatch of query and index languages appears to be the most formidable obstacle to a satisfactory solution of the problem of information retrieval.

---

explain the brain's associative power, its capability of bringing about combinations of terms in form of wordfields and concepts, facilitating comparisons, definitions, classifications and through organizational and sifting processes which lead to evaluations, that is, to productive and creative efforts and finally to sets of entirely original information. It is not merely the number of the memory cells, nor the types and number of circuits and components but the non-linearity, the quality of operation, especially the seemingly unending capability of the conscious mind to comprehend large amounts of most complex information i one single extremely brief thinking process, that characterizes the operation of a productive human brain.

Let us briefly recall first, Trier's word fields and their continuous changes not only within, but also in relation to each other, especially during a period when scientific special areas double in intervals of 8 to 10 years; second, the fields as modified by Quillian, where the type node is described by all the token nodes lying in the same plan, and each token node converted into a type node commands a field of its own with all the corresponding token nodes; and third, a possible combination of both types that would represent a system of continuously changing Quillian-style fields.

As a minimum requirement an "ideal" thesaurus must reflect such a sophisticated, flexible system as to permit identification and correlation of all pertaining terms in and across the different fields in which they may be located. Whether this can be accomplished is still an open question. The ABC method shows certain similarities to the "ideal" in its clusters, its ABC descriptors, and its entire complexity in that the cluster-forming keyword is identified or defined by its ABC descriptors and their constituent terms and phrases, which in turn assemble their own clusters and in this constellation gain substance, meaning and precision.

When we accept Hoenigswald's and Chomsky's[36] theses that the sentence forms the basis of all meaningful communications, the ABC method will have an apparent edge over its competitors unless sentences are incorporated or terminology and phrases are encoded to reconstruct sentence-like combinations, a possibility we will discuss below in a different context.

Our survey of linguistic theories convinced us that:

1. The conventional coordinate-index approach cannot satisfactorily solve the information retrieval problem and that

2. The first-generation ABC method falls also short of coping with the complexities of constructing the desirable multi-dimensional interlinkage of terms, phrases and subject areas, of standardizing the semantics and syntax of the lengthy descriptors, and in particular, of producing the solid foundation on which to develop the progressive automation of all input operations.

Numerous methods have been introduced to improve both types of IR systems. In the thesauri of coordinate indexing systems cross references in limited numbers have been recorded and indices appended that organize the descriptors by classes or groups.

In the first generation ABC dictionary co-occurring terms and phrases, serve the user as guide posts to other clusters and thereby greatly support the retrieval effort. Additional measures

were taken in the second-generation model, not only to simulate more complex language and thinking phenomena, but also to initiate a program for the progressive automation of the analytical input operations.

These techniques have already been subjected to preliminary tests and partly used as standard practice by our analysts and will be explained in more detail subsequently.

k. Focalization Upon Specific Requirements

If an information office is to support the research efforts of scientists and engineers adequately, it must introduce a filter system into its recall mechanism for the sake of effectiveness, economy and utility and must be able to anticipate the particular objectives or operations of its clientele.

In HDL, for example, these objectives comprise the development of sub-systems, devices, components and materials, their properties affecting the performance of the developed hardware, of scientific principles upon which the design of modern equipment can be based, and applicable production and testing methods. Information acquired for HDL must therefore be processed on the basis of relatedness to these particular tasks, missions, and interests. The determination of factual and anticipated relatedness to HDL research and development missions is a major analytical effort within the second-generation ABC development program.

Any new document entered into its information system must not only be identified for its value and overall usefulness but also for the classes and categories to which it should be assigned.

C. THE PROTOTYPE AUTOMATED ABC RETRIEVAL SYSTEM

The results of the manual test performed with the first-generation ABC model had given rise to a number of questions. In particular, the low retrieval effort of our volunteer operators was a weak spot and independent observers were entitled to ask what would happen if the retrieval effort were intensified, when will the law of diminishing returns set in, and at which phase of a steadily continuing retrieval run will the "inverse relationship of relevance and recall" destroy the usefulness and economy of the output.

We acknowledged also the requirement of a new yard stick that would permit fast and valid comparisons of different retrieval runs and different methods.

Neither problem could find a practical solution until the retrieval operation was extended up to a 100 percent recall and the human retrieval operator entirely eliminated and replaced by a completely automated process.

The following paragraphs describe the design, the performance, and the results of such an automated test with particular emphasis on the factors that must affect the already-initiated experiment with an adequately large (second-generation) test collection.

a. Test Collection

The collection for the automated test covers particular aspects of solid state physics and solid state devices and was derived from the materials we had prepared for the experiment of the first-generation ABC model. There were 300 documents selected at random from a total of 3600 (Chart A) as were 50 of a total of 139 related questions.* In this way, we could operate with questions derived from the body of the documents as well as from the general knowledge of the contents of the collection.

b. Processing for Automation

In order to automate the retrieval operations, the descriptors from documents as well as queries were processed by a vector-type method similar to those used by Assario[30],[31] and Salton.[32],[33]

*Of the 50 questions used, thirty-five (Q201 to Q235) had been originally derived from the content of documents in the test collection. The remaining fifteen questions (Q236 to Q250) were formulated with respect to problems of solid state physics and electronics and without inspection of the test set.

Because we utilized two different types of queries, we had to determine the bias possibly introduced by the second group of the specially prepared queries.

The test revealed that they were responsible for a higher deficiency (D=5.00 percent) than those formulated on the basis of a true requirement (with a D of 4.33 percent). We must, however, take into consideration that with the reduction of the samples (to 35 and 15 queries) the confidence level was lowered, and that with the two standard deviations of 0.8 and 1.3 percent overlapping each other, a definitive judgment cannot be made at this time.

Because the amount of information was relatively small the psychometric approach had to be used. The descriptors and queries, each as a complete entity, were related to 59 categories and the degree of relatedness expressed by a scale using the integers 0 through 9. The categories at this time had been pre-established (Chart B) based on missions, subject specialties, and hardware. They were intentionally overlapping because this procedure introduced a certain degree of redundancy which in turn may have been responsible for the matching results that were less dependent upon attitudes and work intensity of the individual evaluators. It is recognized that this subject may require more study and possible changes after the results have been fully analyzed.

The cooperating evaluators, 29 scientists and engineers of the organization consisted of 4 Ph.D, 6 Master, and 19 Bachelor degree holders who were well acquainted with the subject matter. They had the choice of selecting or rejecting the documents and queries to be evaluated. They were instructed to express the degree of relatedness to as many categories as possible (and not only to the important ones) by comparing the content of each paper (or its descriptor) with every single category and by using their best judgment. It was assumed that these very general instructions would prevent the introduction of bias.

No one participating in the test saw the documents themselves. The evaluation was entirely based on the previously established ABC descriptors. In this way we eliminated the bias possibly introduced by the producers of the descriptors (analysts).

Document descriptors were distributed according to the fields of interest and specialization to the scientists. Each descriptor was evaluated by an average 4.2 scientists.

In the retrieval process the computer matched vector-like numbers which represented raw averages over about 4 evaluations of queries and documents, produced the correlation coefficients as a relevance measure for any given query-document pair and generated a printout of these values in descending order to reflect the respective degree of relatedness.

c. The Measuring Tool and its Applications*

Because a large number of different retrieval runs had to be

*In the subsequent paragraphs (p. 26-32) the methodology and the results of the automated prototype test are described for the general reader. The statistically trained documentalist is advised to turn to p. 33.

measured and compared with each other to determine the impact of various parameters and methods and because at a later phase test results from competing retrieval systems would have to be reconciled for analyses and evaluation, we evidently required an independent, flexible, generally applicable and relatively sensitive measuring tool that provided quantitative results from which to draw valid conclusions at a high level of confidence.

The ranked order output furnished the basis for developing such a new tool for evaluation. The minimum number of permutations τ necessary to bring a given output into an arrangement in which all the relevant documents are placed at the top of the lists, was used to express the degree of quality. In order to make this measurement independent of the size and the composition of the collection or the particular preferences of individual requestors we effected normalization by dividing τ by the product of the number of corresponding relevant (r) and nonrelevant (s) documents. This normalized performance measure of a single retrieval run or entire system is therefore represented by the equation D (deficiency)= $\frac{\tau}{r \times s}$. Because it indicates the system's power of placing or diffusing the desirable (pertinent) documents within the ranked-order output, particular parameters such as strength and size of the collection or filters adaptable to peculiar personal requirements (Chart H) must and will be introduced to refine the general, one-digit performance index.

Through this normalization process, the statistical model developed for the evaluation of entire systems as well as for partial or limited retrieval runs has been reduced to five parameters: the recall (ρ); the relevance (p); the cut-off ratio (π); the efficiency or value of the entire collection (α) with regard to a given question or to the average question of a completed test, and the deficiency (D) as previously defined.

The usefulness of the model as an evaluation tool can be explained also in the following manner.

The first four parameters are represented by the equations:*

$$\rho = \frac{RR}{r}; \qquad p = \frac{RR}{RR + NR}$$

$$\pi = \frac{RR + NR}{r + s}; \qquad \alpha = \frac{r}{r + s}$$

* RR= number of relevant items retrieved
RR + NR= number of items withdrawn
r+s= number of items in the collection
r = number of relevant documents in collection
s= number of non-relevant documents in the collection

Because, "p" can be determined without appreciable expenditure and the value of r+s (size of the collection) is always known in every well-administered library, there remain the parameters or parameter elements R= the number of relevant or responsive titles in the collection, and D= the deficiency which may be difficult to estimate.

The dependence of the relevance=recall curve (Fig. 4) on the value αK (that is on r and on D) indicates that without a predetermined r and without a D pre-established for the system we cannot calculate the optimum cut-off for a manual system. For an automated system (with a ranked-order output), however, we turn the defect of the interdependence of relevance and recall into a major advantage by using the curves to predict on the basis of p and π the value of D and subsequently the number of relevant documents in the collection and the limit of an efficient retrieval run.**

Although D has its definition and its values derived from the results of a ranked-order retrieval process, there are indications that it might prove to be a useful evaluator applicable to all, including manual, retrieval methods.*** When we use this method for an evaluation of the manual test of the first generation model happens to be 0.3 percent, an apparently surprisingly low value that can be easily explained. People using the ABC dictionary for their own requirements can accurately determine those descriptors that immediately lead to documents. This advantage is, however, partly offset by their unwillingness or disinclination to persevere in their effort, to obtain an optimum recall. From the calculated D and the appropriate relevance recall curve of the model (Fig. 4 and 5), we can estimate that if the operators had continued their manual search to retrieve an average of 9-10 documents per run, the recall would have risen to 75 percent, but the corresponding relevance ratio declined to about 70 percent. We must therefore draw the following conclusions:

**The normalized relevance-recall curves form a family of hyperbolas which with increasing deficiency deteriorate into approximations of straight lines. Provided the statistical level of confidence is sufficently high, retrieval data permit evaluations of the strength of the collection (with respect to the particular query) as well as of the quality of the system. The constant inverse relevance-recall relationship expressed by an approximately straight line, for example, is generally an indication of a low efficiency (high D) system.

***For a derivation of K and therefore also of D from most general considerations. See below p. 37-38.

(1) The law of diminishing returns prevails in every system if D>0.
(2) The rapid deterioration of relevance along the curve is more pronounced with increasing D. For a low deficiency (D) the relevance is high and almost constant up to a high level recall ratio. For high deficiencies the curve comes close to a straight line.

The widely differing D values for the automatically obtained retrieval results and for those of the manual (first-generation ABC model) test warrant some explanations.

For the vector test the input had been processed by a combined human effort, the psychometric method. To the extent this human effort, in addition to the one spent on the generation of the ABC descriptors, has affected the quality of the organized collection, it contributed to a lowering of the value D. If we could at this time isolate this deteriorating element for example by automating the input analyses,* the D derived from the vector test should turn out to be better rather than worse.

How much better, we hope to determine in one of our subsequent test runs when a vector type organization will be produced not by the psychometric approach but entirely by mechanical means. It is for this purpose that we require the larger test collection. We shall use this collection also for an additional purpose, the evaluation of the model's capacity of estimating the optimum cut off for particular automated retrieval runs.

d. Statistical Problems

The performance measure was also made to relate the results of the mechanical retrieval to the number of relevant documents included in the collection.

Such a provision is obvious when all or none of the collection's documents are pertinent to the query. In these extreme cases the results do not measure the quality of the system. In fact, no system can be evaluated without reference to the strength of the holdings, i.e. to the parameter α.

According to statistical principles the accuracy of the performance measure will improve with the increasing size of collection and samples.

The average deviation of the observed D from average D (4.8 percent in the mechanized model) can be determined by the formula $\Delta D \sim r^{-1/2}$ and the distribution of the relevant documents among the non-relevant ones according to their ranked position on the probability

*See Appendix III.

curve (dashed line in Figure 2) rather than according to interdependence (e.g. spatial organization) of documents whereby the true situation would only be complicated and obscured.

To demonstrate the validity of this assumption we plotted the deficiencies (D_1) of all returns from one run vs. the corresponding numbers of available documents relevant to the particular query. This diagram (Fig. 8) corroborates the expected random distribution in accordance with the standard deviation formula.

The size of the (300 title) collection and the number of questions (50) were the determining factors for a mean error that was just small enough to permit the evaluation of different retrieval methods and to make estimates with an acceptable, high confidence level.

e. Elements Tested

In different test runs we determined: (a) the best matching formula out of four when we used 10-scale evaluations for the construction of vectors; (b) the optimum method of reducing the 10-scale evaluation to a 2-value scale; (Figure 11); (c) the contributions made by the individual categories; (d) the significance of the number of categories, that is, of the dimensions of the vector spaces, and (e) the influence of number and quality (knowledge, ability and experience) of evaluators we employed to determine the vector components.

f. Impact of the Number of Categories in a Given System

Thirteen test runs proved the deficiency D to be inversely proportional to the number of categories in the system (Fig. 13). Translated into practice it means that doubling the number of categories in a given system could compensate for a reduction of the 10-value to a 2-value scale(See fig. 11); or a 180-category system to be used for the entire HDL collection (this is tripling the number of categories in the system of the prototype collection) and a simultaneous reduction of the 10-scale to the programmed 3-scale evaluation may result in a better deficiency D than the one observed with the prototype test.

The improvement would approximate 55 percent, if the impact of the larger number of subjects covered should turn out not to be a deteriorating factor.

We had no intention of improving the organization of the small test collection although we were aware that improvements of

various elements could be obtained by statistical means. We were aware that the collection was relatively small, and that any decrease of the deficiency D might relate only to this particular sample of documents and would not necessarily apply to another, larger collection, even though it was organized by the same principle. We were merely guided by the requirement of establishing the processing methods for the new and larger test collection when we tried different ways of raising the quality of the model.

For assessing the contribution any one specific category had made to the test results, we performed six tests by dropping each time one particular category from the document and query vectors. The results showed an impact ranging widely from an improvement of 10% per category down to a deterioration of .1% (Fig. 13). The last-mentioned result is an indication that the system can be improved if certain categories, especially those which are subject to a great variety of interpretations, will be excluded.

Another approach consisted of calculations of "importance values" for all categories. In the same process we determined also the absence of obvious correlations between frequency of utilization and "importance values" (Fig. 14). A low "importance value" may point to the lack of an adequate representation in a particular category or a deficient treatment of the subject area by some of the evaluators. Nevertheless, we used the "importance values" as weight factors to improve the results of the matching process. In this way the quality of the system was slightly enhanced as demonstrated in particular when the values calculated by way of dropping six individual categories were inserted (Fig. 15).

If this method were extended to all of the 59 categories we could expect an oberall gain of 25-30 per cent. In future tests weight factors can also be obtained through user feedback that might coincide with a requested reaction to a service of disseminating selected information.

The categories we used were overlapping. Those denoting such general categories as design, development, were useless; some worsened the results because they were checked either too often or interpreted differently by different evaluators. On the other hand, categories representing clearly defined subjects e.g. disciplines or hardware, contributed greatly to good retrieval performance.

Although 180 categories will be initially applied to organize the new and more complex test collection of the 2nd generation ABC model, this number can probably be reduced by factorization without a deterioration of the results.

g. On-Line Retrieval

In a small experiment we demonstrated the capability of the system[34] of having the individual scientist and engineer perform his own retrieval operations directly from the laboratory.[35] In connection with this demonstration a small collection of 400 documents (processed according to ABC-standards) was programmed for on-line retrieval. The information was stored on the drum memory of a computer located in Los Angeles. From a terminal in HDL the experimenter requested the transmission of ABC descriptors by feeding in the keywords which he selected for the importance to his query from a prototype category term dictionary. The computer responded by transmitting: first, statistical information; second, the full texts of the appropriate ABC descriptors; and third, the full titles with shelf numbers after the codes of the selected descriptors had been typed in by the requestor.

The advantage of displaying the ABC descriptors first is two-fold. The investigator can identify and select appropriate documents on the basis of standardized descriptors prepared by research analysts who are thoroughly familiar with the installation's missions and requirements; and he can continue his search by keying in terms and phrases that he encounters in the displayed descriptions and in this way can exploit the entire collection without regard to the category he used to enter the systems. The successful transmission gave ample proof that the system is well suited for retrieval of ABC-processed information from a nearby computer system. It appears that the COLEX[36] program (developed for DIA) could be adapted to automate the ABC system for direct retrieval from our laboratories and for permitting an immediate confrontation of the investigator with the organized collection. In case of a successful application cross sections of up to 10 terms as well as personal or corporate author names could initiate the retrieval of the corresponding ABC descriptors and finally lead to the complete bibliographic information. There would be no longer a need for printing the complete ABC dictionary.